Virology 398 (2010) 135–147

Contents lists available at ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/yviro

A jumbo phage infecting the phytopathogen *Ralstonia solanacearum* defines a new lineage of the Myoviridae family

Takashi Yamada [a,*], Souichi Satoh [a], Hiroki Ishikawa [a], Akiko Fujiwara [a], Takeru Kawasaki [a], Makoto Fujie [a], Hiroyuki Ogata [b]

[a] *Department of Molecular Biotechnology, Graduate School of Advanced Sciences of Matter, Hiroshima University, 1-3-1 Kagamiyama, Higashi-Hiroshima 739-8530, Japan*
[b] *Information Génomique et Structurale, CNRS-UPR2589, Institut de Microbiologie de la Méditerranée, Aix-Marseille Université, 163 Avenue de Luminy, Case 934, 13288 Marseille Cedex 9, France*

ARTICLE INFO

Article history:
Received 2 September 2009
Returned to author for revision
31 October 2009
Accepted 25 November 2009
Available online 24 December 2009

Keywords:
Jumbo bacteriophage
Ralstonia solanacearum
Genomic analysis
Proteomic analysis
Gene expression

ABSTRACT

ϕRSL1 is a jumbo myovirus stably and lytically infecting the phytopathogenic bacterium *Ralstonia solanacearum*. In this study, we investigate the infection cycle of ϕRSL1 and provide a genomic, proteomic and transcriptomic view of this phage. Its 231-kbp genome sequence showed many genes lacking detectable homologs in the current databases and was vastly different from previously studied phage genomes. In addition to these orphan proteins, ϕRSL1 was found to encode several enzymes that are unique among known viruses. These include enzymes for the salvage pathway of NAD^+ and for the biosynthetic pathways of lipid, carbohydrate and homospermidine. A chitinase-like protein was found to be a potential lysis enzyme. Our proteomics analysis suggests that ϕRSL1 virions contain at least 25 distinct proteins. We identified six of them including a tail sheath protein and a topoisomerase IB by N-terminal sequencing. Based on a DNA microarray analysis, we identified two transcription patterns.

Introduction

Bacteriophages constitute a majority of biological organisms on the Earth and have crucial influences on the evolution of bacteria (Ashelford et al., 2003; Hendrix 2002, Suttle, 2005; Wommack and Colwell, 2000). The double-stranded DNA (dsDNA)-containing tailed phages (the Caudoviridales) represent the most numerous, most widespread and probably the oldest group of bacteriophages (Hendrix, 1999). Three families (i.e. Myoviridae, Podoviridae, and Siphoviridae) belong to the order Caudovirales, and approximately 25% of the members of this order are myoviruses (Ackermann, 2003). Myoviridae phages classically represented by T4-like phages have a contractile tail sheath and infect a broad range of bacterial hosts. From analyses of T4-like phage genomes, it has been suggested that myovirus genomes are mosaic of conserved core genes, which include structural genes encoding head and tail proteins and enzyme genes for DNA and nucleotide processing (Filee et al., 2006), and the remaining accessory non-core genes, which are not conserved across species and correspond to the vast majority of the huge gene pool of phages. The functions of the non-core genes are mostly unknown, although it is commonly assumed that they provide a selective benefit to phages (Hendrix, 2009).

Several myoviruses are known to have a large genome over 200 kbp, and are designated as "jumbo phages" (Hendrix, 2009). These include *Pseudomonas aeruginosa* phage ϕKZ (280 kbp, Mesyanzhinov et al., 2002) and EL (211 kbp, Hertveldt et al., 2005), *Vibrio parahaemolyticus* phage KVP40 (386 kbp, Miller et al., 2003a), *Stenotrophomonas maltophilia* phage ϕSMA5 (250 kbp, Chang et al., 2005), and *Yersinia enterocolitica* phage R1-37 (270 kbp, Kiljunen et al., 2005). Jumbo phages were also reported for *Sinorhizobium meliloti* (phage N3, 207 kbp; Martin and Long, 1984) and *Bacillus megaterium* (phage G, 600 kbp; Sun and Serwer, 1997). Recently, Yamada et al. (2007) isolated a new lytic jumbo phage (ϕRSL1) with a 240-kbp dsDNA genome, which efficiently and stably infects a wide-panel of *Ralstonia solanacearum* strains. The ϕRSL1 particle has an icosahedral head of 150 nm in diameter and a contractile tail of 138 nm in length and 22.5 nm in diameter (Fig. 1).

R. solanacearum is a soil-borne Gram-negative bacterium (Beta-proteobacterium) known to be the causative agent of bacterial wilt of many economically important crops (Yabuuchi et al., 1995; Hayward, 2000). The bacterium has an unusually wide host range covering over 200 plant species belonging to more than 50 botanical families, and shows great phenotypic and genotypic diversity between strains (Hayward, 2000). To investigate potential roles of phages on the biological, physiological, and pathological variations of *R. solanacearum*, we determined the genome sequence of ϕRSL1, characterized phage gene expression patterns during infection, and performed a proteomic analysis of virions.

* Corresponding author. Fax: +81 82 424 7752.
E-mail address: tayamad@hiroshima-u.ac.jp (T. Yamada).

0042-6822/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.virol.2009.11.043

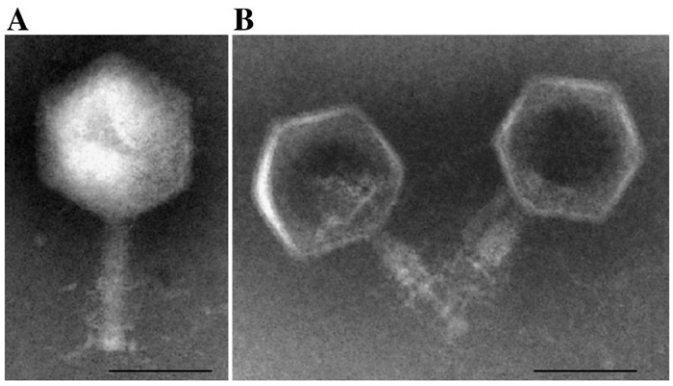

Fig. 1. Electron micrographs of ϕRSL1 particles (A and B). Empty capsids with a contracted tail are shown in B. Phage particles were stained with 1% sodium phosphotungstate. Scale bar represents 100 nm.

Results and discussion

Growth cycle of the jumbo phage ϕRSL1

ϕRSL1 could form clear plaques in top gel containing 0.45–0.5% agar but formed no visible plaques in gel of a higher agar concentration (0.7–0.8%) used in a standard plaque assay. The infection cycle of ϕRSL1 was characterized by a single-step growth experiment with strain M4S as a host. As shown in Fig. 2, the eclipse phase period was found to be ~90 min, and the latent period ~150 min. These were followed by a rise period of 90 min. The average burst size of ϕRSL1 was 80–90 pfu per infected cell. The DNA replication started around 75 min post infection (p.i.) as revealed by DNA dot blot hybridization of total bacterial DNA with a ϕRSL1-probe (data not shown). The use of other *R. solanacearum* strains, such as MAFF301558 and MAFF730138, led to essentially the same results. Lytic cycle and the lack of lysogeny of ϕRSL1 was supported by genomic Southern blot hybridization assays in which 15 host strains (of different races and/or biovars including ϕRSL1-sensitive and -resistant ones) showed no hybridization signal with a labeled ϕRSL1 DNA probe (Yamada et al., 2007). One characteristic feature of ϕRSL1 infection was a lasting host killing effect. In liquid cultures, no discernible growth of resistant bacterial cells was observed until at least 100 h p.i.

General features of the genome and the abundance of ORFans

ϕRSL1 shotgun sequences were assembled into a closed, circular genome of 231,255 bp, which was slightly smaller than the previous estimate (i.e. 240 kbp) by pulsed-field gel electrophoresis (Yamada et al., 2007). The small difference suggests that ϕRSL1 phage particle may contain a linear, circularly permuted, and terminally redundant DNA genome like T4 (Miller et al., 2003b). The average G + C content of the ϕRSL1 genome was 58.2%, being substantially lower than those of the 3.7-Mbp large replicon (67.04%) and 2.1-Mbp small replicon (66.86%) of *R. solanacearum* GMI1000 (Salanoubat et al., 2002). In the phage genome, we identified 343 open reading frames [ORFs; from 40 to 1586 amino acid residues (aa)] and 3 tRNAgenes (including one pseudogene). Few paralogs or repetitive sequences (>50 bp) were found in the genome. According to the orientation of ORFs, we identified four major genomic regions (Fig. 3 and Supplementary Table S1). The largest region I (114.5 kbp) encodes 193 ORFs (ORF001–ORF146 and ORF296–ORF343) mostly (186 ORFs) located in a clockwise orientation. This region has an average G + C content of 58.3%. Region II (15.8 kbp, G + C = 57.4%) encodes 24 ORFs (ORF147–ORF170) all in a counterclockwise orientation. Region III (27.9 kbp, G + C = 57.4%) encodes 23 ORFs (ORF171–ORF193; 19 clockwise and 4 counterclockwise). The last and the second largest region IV (73.1 kbp, G + C = 58.4%) encodes 102 ORFs (ORF194–ORF295) mostly (100 ORFs) located in a counterclockwise orientation. No repeated sequences or transposon-related sequences were found near the junctions between these four genomic regions, except for ORF296 (at the IV/I junction) corresponding to a putative recombination endonuclease. ORFs were tightly organized in general with little intergenic space, and in many cases, the stop codon of an ORF was found to overlap the start codon of the following ORF.

ϕRSL1 ORFs showed no detectable similarity at the nucleotide level in other viruses or in cellular organisms, and significant sequence similarities, if any, were detected only at the protein sequence level. The proportion of ORFans (i.e. genes lacking detectable homologs in the current databases) was very high. Of the 343 RSL1 ORFs, 251 ORFs (73%) showed no significant sequence similarity in the publicly available databases (*E*-value<0.001). However, due to the use of different methods in gene prediction and homology search, this ORFan proportion was not directly comparable to those previously published for other phages. Therefore, we performed BLAST searches for available phage genomes by restricting our analysis to long ORFs (≥150 aa), for which the effect of false gene identification would be negligible. The result (Fig. 4C) clearly indicates that the ORFan proportion of ϕRSL1 was the highest among the jambo phages with genomes larger than 200 kbp. Of 167 long ORFs (≥150 aa), 96 ORFs (57%) had no significant match in the UniProt database.

Of all the 343 ORFs, 83 had homologs in UniProt or in the NCBI environmental sequence collection (GenBank/gbenv). Of these, 53 ORFs (15.5%) showed best sequence similarities (*E*-value<0.001) in bacteria, 10 ORFs (2.9%) in viruses/plasmids, 1 ORF (0.3%) in eukaryotes, 1 ORF (0.3%) in archaea, and 18 ORFs (5.3%) in environmental sequences (Fig. 4A). Taxonomic distribution of the best homologs is shown in Fig. 4B for those ORFs that showed most similar sequences in the UniProt database. These include Betaproteobacteria (12 best homologs, 7 of which are from *Burkholderia* spp. and 3 are from *Ralstonia* spp.), Alphaproteobacteria (10 homologs including 4 from *Rhizobiales* and 3 from *Rhodobacteriales*), Gammaproteobacteria (9 homologs), Epsilonproteobacteria (3 homologs), Firmicutes (3 homologs), Verrucomicrobia (3 homologs), Actinobacteria (2 homologs), Cyanobacteria (2 homologs), Deltaproteobacteria (2 homologs), and other bacteria (7 homologs). It is notable that only a few ϕRSL1 ORFs showed most similar sequences in *Ralstonia* spp., given that the genomic sequences available for *R. solanacearum* strain GMI1000 (race 1, biovar 3, and phylotype I; Salanoubat et al., 2002) and strain UW551 (race 3, biovar 2, and phylotype II; Gabriel et al., 2006). Viruses/plasmids best hits include 5 homologs in myoviruses, 3 in siphovirus, 1 in eukaryotic virus (Mimivirus) and 1 in plasmid. These results suggest that ϕRSL1 may have access to the gene pools of

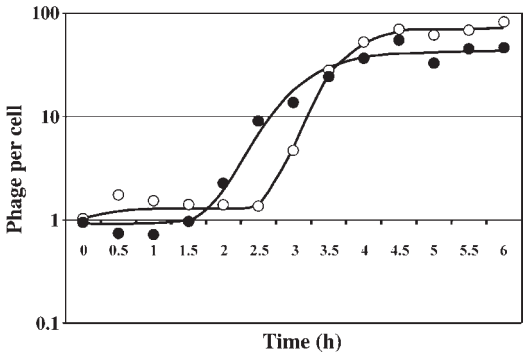

Fig. 2. Single-step growth curve of ϕRSL1 growing on *R. solanacearum* M4S. Shown are the PFUs per infected cell in cultures at different time post infection. Samples were taken at 30-min intervals up to 6 h and immediately diluted with (filled circle) or without (open circle) chloroform treatment, and the titers were determined by the double-layered agar plate method. Eclipse and latent periods were ~90 min and ~150 min, respectively.

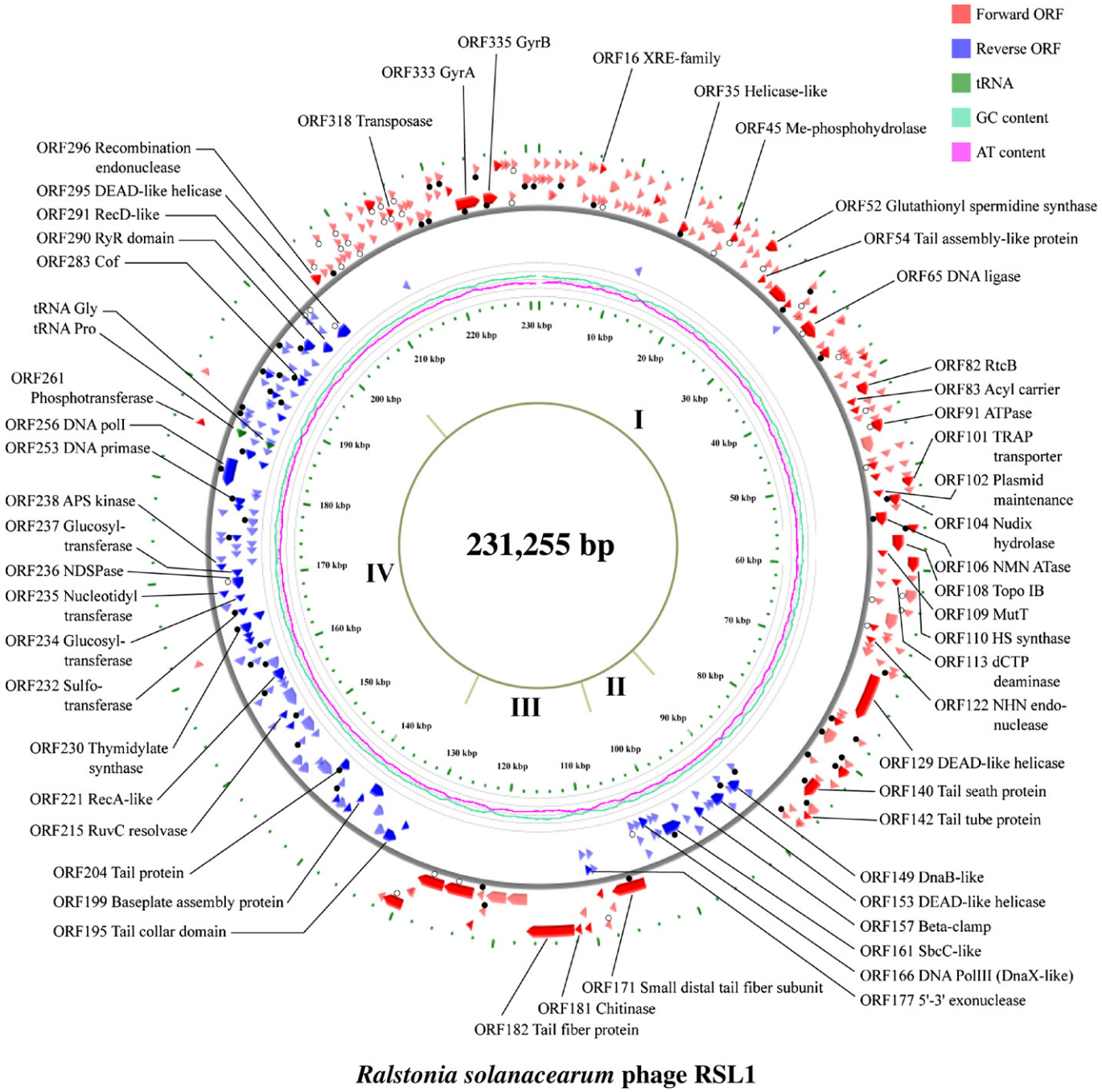

Fig. 3. Organization of the ϕRSL1 genome. Putative ORFs are indicated by red arrowheads (in clockwise orientation) and blue arrowheads (in counterclockwise orientation). Pale blue and pink lines show the G+C and A+T contents, respectively. Genomic regions I, II, III, and IV are indicated along the inner circle. Open and filled circles indicate ORFs expressed in early-intermediate stages and in intermediate–late stages of RSL1 infection, respectively.

largely different families of bacteria and viruses. We could assign putative functions to 47 ORFs through the examination of homology search results. These are distributed across several functional categories including DNA replication and processing, phage structural components, nucleotide metabolism, and a variety of other unique metabolisms (Table 1 and Supplementary Table S1).

DNA replication/processing genes and DNA polymerase phylogeny

Many of the core genes conserved in various phages are those involved in DNA replication, recombination, repair and packaging. In the genome of ϕRSL1, we found homologs for DNA polymerase I (family A DNA polymerase, PolA; ORF256), DNA polymerase III (ORF166), DNA ligase (ORF065), DNA primase (ORF253), DNA topoisomerase IIA (ORF333 and ORF335), clamp (ORF157), RNaseH (ORF177), DnaB helicase (ORF149), DEAD-like helicase (ORF129, ORF153, and ORF295). To examine the evolutionary relationship between ϕRSL1 and other phages, we performed phylogenetic analyses of DNA polymerase I homologs (PolA), which often serve as a phylogenetic marker. The ϕRSL1 PolA sequence was found to be placed far away from other myovirus homologs (Supplementary Fig. S1). It was grouped with homologs from *Enterobacteria* phage T5 and *Pseudomonas* phage Pap2, but this grouping was not statistically supported. Additional phylogenetic analyses with homologs from

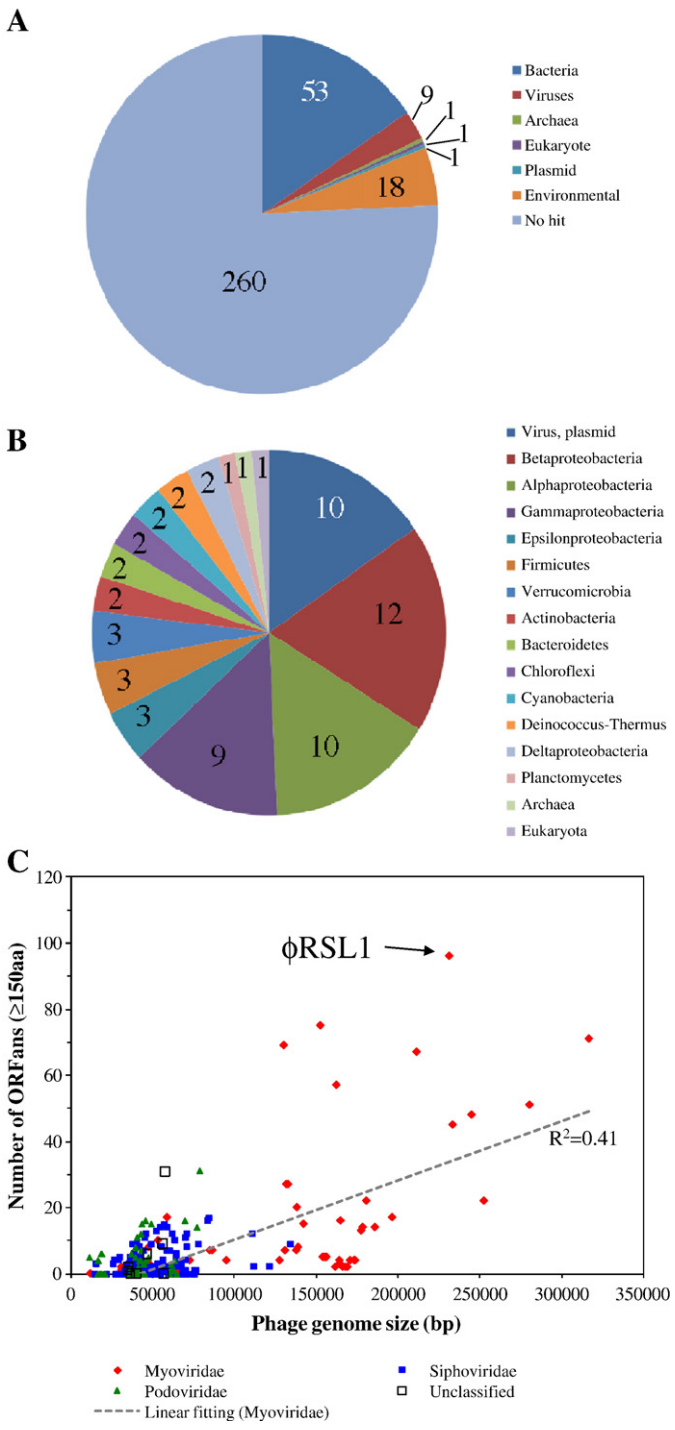

Fig. 4. Taxonomy of best Blastp hits of ϕRSL1 ORFs. (A) Best hits to UniProt/NCBI-Env data. Among 343 ϕRSL1 ORFs, 83 hits (*E*-value<0.001) were detected and classified. 260 ORFs (75.8%) showed no hits (ORFans). (B) Sixty-five best hits to UniProt are further classified. (C) Number of unique ORFs in the genome vs. genome size of various phages sequenced to date. Unique ORFs here are ORFs (>150 aa) lacking Blastp hits (*E*-value<0.001) in the UniProt sequence database. Very close hits (percent identity >98% and alignment coverage >90%), present in most case self-hits, were discarded from the analysis.

diverse bacterial species (including *R. solanacearum*) further confirmed the lack of specific phylogenetic affinity between ϕRSL1 PolA and its homologs (data not shown). We also noted that the presence of a PolA in ϕRSL1 is unique among jumbo phages (>200 kbp). Other jumbo phages encode either family B DNA polymerase homologs or no obvious DNA polymerase homologs [e.g. *Pseudomonas* phages KZ (Mesyanzhinov et al., 2002) and EL (Hertveldt et al., 2005)]; ϕRSL1 genome is thus the largest among the PolA-encoding phage genomes (Supplementary Fig. S1). ϕRSL1 ORF065 showed the highest similarity to the NAD^+-dependent DNA ligase of Mimivirus (MIMI_R303, YP_142657, *E*-value of 1e-34). Mimivirus infects *Acanthamoeba* spp. and is the largest virus known to date with a genome of 1.2 Mbp containing 911 protein containing genes (Raoult et al., 2004). Other ORFs possibly involved in DNA recombination/repair include a helicase (ORF035), a topoisomerase IB (ORF108), a SbcC-ATPase involved in DNA repair (ORF161), a RuvC resolvase (ORF215), a RecA (ORF221), a RecD exodeoxyribonuclease V (ORF291), and a recombination endonuclease (ORF296). In T4, many DNA replication/processing enzymes are encoded within a single gene cluster of ~20 kbp. In the ϕRSL1 genome, many of the ORFs of this category are located in region II and IV (i.e., counterclockwise gene clusters, Fig. 3) and several others are dispersed within region I. Region III was found to contain none of them (Fig. 3).

Proteomics of virion

Homology searches against multiple sequence/profile databases detected no clear homolog for the major capsid proteins in the genome of ϕRSL1. In an attempt to identify the major capsid protein and other structural components, we performed a proteomic analysis of ϕRSL1 virions by SDS-polyacrylamide gel electrophoresis (PAGE). At least 25 proteins ranging from 13 kDa to 160 kDa were separated in the gel (Fig. 5). Six well separated bands with molecular masses of 17.0, 18.5, 24.7, 33.0, 55.0, and 65.0 kDa were recovered from the gel and subjected to N-terminal sequencing (Fig. 5). According to the determined N-terminal sequences, they were found to correspond to ORF133, ORF121, ORF135, ORF136, ORF108, and ORF140, respectively. Posttranslational processing by proteolysis is necessary for the morphogenesis of many phages (Black et al., 1994). Consistently, several N-terminal residues in the conceptual translation of these ORFs were missing in the experimentally determined N-terminal sequences: 31 aa from ORF133, 1 aa from ORF121, 2 aa from ORF135, 95 aa from ORF136, 14 aa from ORF108, and 11 aa from ORF140. No significant sequence homology was found for these ORFs except for ORF108 (A0QTB9_MYCS2, *Mycobacterium smegmatis* topoisomerase IB, *E*-value of 5e-06) and ORF140 (NP_944106, tail sheath protein (gp18) of bacteriophage Aeh1, *E*-value of 8e-30). The 33.0-kDa protein band (ORF136) represents the most abundant protein, thus likely corresponding to the major capsid protein. ORF136 (362 aa) showed no significant database homolog; its BLAST best hit was the putative sigma54 specific transcriptional regulator from *Clostridium* with an *E*-value of 0.8, which was not considered significant. Interestingly, this 33.0-kDa protein appears to be produced by a cleavage after a Gly residue (aa position 95 of ORF136), which is also the conserved residue in the proteolytic cleavage site of phage T4 head protein (Black et al., 1994). In support of this prediction that ORF136 corresponds to the major capsid protein, we found a marginal homology (*Z* score 3.76; 90% confidence) between ORF136 sequence (after removing the N-terminal 95 residues) and a putative capsid protein of *Escherichia coli* CFT073 prophage (PDB 3BQW) with the use of a sequence-structure homology recognition method (FUGUE, http://tardis.nibio.go.jp/fugue/). The 6 ORFs identified by our proteomic analysis and ORF142 encoding tail tube protein were found to be clustered in a clockwise orientation in the genome (55 kbp–90 kbp; Fig. 3). Other putative structural proteins are located in different parts of the genome: ORF054 encoding phage tail assembly protein is located in a clockwise orientation at a position around 26 kbp, and four ORFs (ORF194 for tail fiber assembly protein, ORF195 for tail fiber, ORF199 for base plate assembly protein), and ORF204 for a tail protein are located in a counterclockwise orientation at positions from 130 kbp to 145 kbp (Fig. 3). Finally, ORF171 and ORF182 were similar to tail fiber proteins from T4 (gp36, Q19CF5, *E*-

Table 1
Features of ϕRSL1 ORFs, gene products, and functional assignments.

Functions	CDS	Strand	CDS start position	CDS stop position	Length (aa)	Amino acid sequence identity/ similarity to best homologs	Accession no.	BLAST score (*E*-value)
DNA replication, recombination, and repair enzymes	RSL1_ORF035	+	15643	16605	320	Helicase superfamily c-terminal domain like	cd00079	3.00E-04
	RSL1_ORF065	+	32521	34398	625	NAD-dependent DNA ligase [EC:6.5.1.2]	COG0272	9.00E-35
	RSL1_ORF108	+	56561	582340	559	Topoisomerase IB	cd00659	3.00E-17
	RSL1_ORF129	+	70925	75592	1555	DEAD-like helicases superfamily	COG0553	8.00E-12
	RSL1_ORF149	−	89031	90374	447	Replicative DNA helicase (DnaB like)	COG0305	1.00E-15
	RSL1_ORF153	−	91500	93029	509	DEAD-like helicases superfamily	COG1061	1.00E-31
	RSL1_ORF157	−	94443	95510	355	Beta clamp domain-like	cd00140	5.00E-05
	RSL1_ORF161	−	97707	99971	754	SbcC (ATPase involved in DNA repair) like	COG0419	2.00E-07
	RSL1_ORF166	−	101309	102397	362	DNA polymerase III gamma/tau subunit-like (DnaX-like)	COG2812	2.00E-30
	RSL1_ORF177	−	109438	110190	250	5′–3′ exonuclease	cd00008	5.00E-32
	RSL1_ORF215	−	152058	152660	200	RuvC_resolvase like	cd00529	8.00E-04
	RSL1_ORF221	−	156275	157720	481	Rec A-like recombinase	COG0468	2.00E-10
	RSL1_ORF253	−	178266	179303	345	DNA primase (DnaG-like)	COG0358	3.00E-08
	RSL1_ORF256	−	180391	183543	1050	DNA polymerase I [EC:2.7.7.7]	cd06444	1.00E-29
	RSL1_ORF291	−	199985	201238	417	RecD (endodeoxyribonuclease V) like	COG0507	5.00E-22
	RSL1_ORF295	−	202619	204433	604	DEAD-like helicases superfamily	COG1061	7.00E-04
	RSL1_ORF296	+	204970	205983	337	Recombination endonuclease like	Q06ER8	2.00E-05
	RSL1_ORF333	+	222295	224943	882	DNA gyrase subunit B	COG0187	3.00E-74
	RSL1_ORF335	+	225283	226827	514	DNA gyrase subunit A	smart00434	3.00E-59
Phage structural proteins	RSL1_ORF054	+	25897	26433	178	Appr-1-pase family/phage tail assembly-like protein	cd02900	2.00E-19
	RSL1_ORF140	+	83105	85051	648	Tail sheath protein	pfam04984	1.00E-29
	RSL1_ORF142	+	86151	86660	169	Tail tube protein	pfam06841	2.00E-06
	RSL1_ORF171	+	104083	107754	1223	Small distal tail fiber subunit	YP_142853	7.00E-10
	RSL1_ORF182	+	112143	116903	1586	Unknown (weak matches to tail fiber proteins)	YP_0012857	7.00E-08
	RSL1_ORF194	−	131919	132695	258	Tail fiber assembly-like protein	YP_717766	9.00E-06
	RSL1_ORF195	−	132703	134199	498	Phage tail collar domain	pfam07484	1.00E-13
	RSL1_ORF199	−	138378	138830	150	Baseplate assembly protein W-like	PHA00415	6.00E-07
	RSL1_ORF204	−	141995	143338	447	Phage-related tail protein	A9ABJ4	3.00E-28
Nucleotide metabolism	RSL1_ORF113	+	61145	61582	145	Deoxycytidylate deaminase domain like	cd01286	4.00E-26
	RSL1_ORF230	−	163104	164348	414	Thymidylate synthase	cd00351	3.00E-29
NAD salvage pathway	RSL1_ORF104	+	52511	53626	371	Bifunctional NMN adenylyltransferase [EC:2.7.7.1]/Nudix hydrolase [EC:3.6.1]	PRK05379	1.00E-94
	RSL1_ORF106	+	54148	55566	472	Nicotinate phosphoribosyltransferase	PRK09198	E-152
	RSL1_ORF109	+	58252	58755	167	MutT/NUDIX hydrolase family protein	pfam00293	8.00E-09
Lipid metabolism	RSL1_ORF083	+	42124	42387	87	Acyl carrier protein	CHL00124	1.00E-12
	RSL1_ORF261	−	186212	186385	57	CDP-diacylglycerol-glycerol-3-phosphotransferase	Q4UV59	1.00E-05
Sugar and polysaccharide metabolism	RSL1_ORF232	−	165261	166118	285	Sulfotransferase-like	pfam03567	3.00E-06
	RSL1_ORF234	−	166983	167750	255	Glucosyltransferase	Q5DN83	1.00E-08
	RSL1_ORF235	−	167743	168465	240	Nucleotidyl transferase/lipopolysaccharide biosynthesis protein	COG1208	2.00E-09
							Q4HLF0	2.00E-14
	RSL1_ORF236	−	168468	169901	477	Nucleoside-diphosphate-sugar pyrophosphorylase-like/Capsular polysaccharide biosynthesis protein	Q8VW65	2.00E-50
	RSL1_ORF237	−	169902	170801	299	Glucosyltransferase	Q5DN83	3.00E-12
	RSL1_ORF238	−	170803	171498	231	Adenosine 5′-phosphosulfate kinase	COG0529	8.00E-13
Miscellaneous	RSL1_ORF122	+	67489	67806	105	HNH endonuclease	cd00085	7.00E-05
	RSL1_ORF332	+	221741	222283	180	HNH endonuclease with an additional AP2 domain	PHA00280	8.00E-06
	RSL1_ORF318	+	215444	215740	98	Transposase	pfam01527	8.00E-05
	RSL1_ORF181	+	111450	112127	225	Chitinase like	COG3179	1.00E-26
	RSL1_ORF016	+	6237	6590	117	Helix–turn–helix XRE-family like protein	cd00093	2.00E-04
	RSL1_ORF045	+	20864	21676	270	Metal dependent phosphohydrolases like	cd00077	7.00E-04
	RSL1_ORF052	+	24276	25397	373	Glutathionylspermidine synthase [EC:6.3.1.8]	pfam03738	E-115
	RSL1_ORF110	+	58761	60203	480	Homospermidine synthase [EC:2.5.1.44]	pfam06408	E-140
	RSL1_ORF082	+	40868	42121	417	Unknown, uncharacterized conserved protein (RtcB)	COG1690	E-104
	RSL1_ORF091	+	44837	46096	419	ATPase (AAA superfamily)	COG0714	1.00E-12
	RSL1_ORF101	+	51051	52001	316	TRAP transporter solute receptor like	COG2358	4.00E-12
	RSL1_ORF102	+	52129	52299	56	Plasmid maintenance system antidote protein like	B5S671	4.00E-06
	RSL1_ORF283	−	195482	195826	114	cof, HAD family hydrolase	COG0561	6.00E-04
	RSL1_ORF290	−	198606	199964	452	RyR domain	pfam02026	1.00E-06
tRNA	RSL1_RNA1	−	186298	186376		tRNA-Gly (anticodon:TCC)		
	RSL1_RNA2	−	186387	186459		tRNA (Pseudogene)		
	RSL1_RNA3	−	186526	186606		tRNA (anticodon:TGG)		

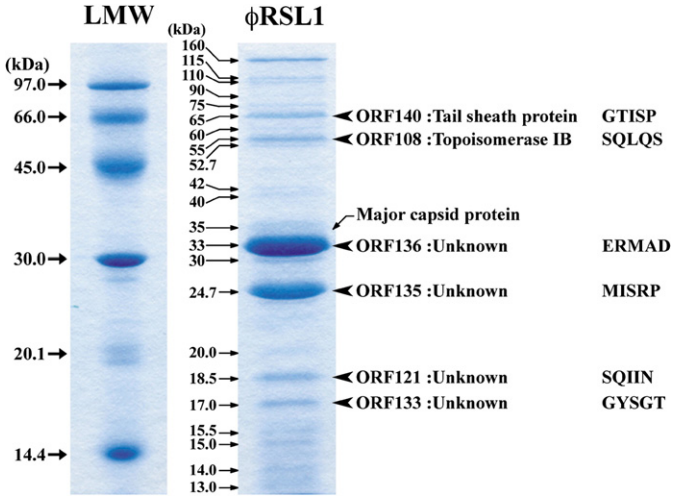

Fig. 5. Identification of ϕRSL1 virion proteins. Proteins from purified ϕRSL1 particles were separated by SDS-PAGE (12% acrylamide gel) and stained with Coomassie blue. The molecular size of each marker protein (an Amersham LMW gel filtration kit) is indicated on the left. The determined N-terminal amino acid sequences of six protein bands (corresponding to ORF108, ORF121, ORF133, ORF135, ORF136, and ORF140) are indicated on the right with ORF IDs. Estimated molecular size for each ϕRSL1 virion proteins is indicated in the middle.

value 1e-07) and from *Phormidium* phage (YP_0012857, *E*-value 7e-08), respectively.

Predicted genes for other metabolic functions

Nucleotide metabolism

T4 and T4-like phages have many enzyme genes for the acquisition of nucleotides used to replicate genomic DNA. Phage-encoded enzymes function to fill the deoxynucleoside triphosphate pool by acting on deoxynucleoside monophosphates from host DNA degradation and on nucleoside diphosphates from host mRNA decay (Greenberg et al., 1994). ϕRSL1 has a thymidylate synthase homolog (ThyA, EC 2.1.1.45, ORF230) and a deoxycytidine triphosphate deaminase homolog (EC 3.5.4.13, ORF113). Thymidylate synthase is involved in an important step of the *de novo* dTMP biosynthesis where it catalyzes the methylation of dUMP to generate dTMP. ORF230 showed the highest amino acid sequence similarity to gp17 of the *P. aeruginosa* siphovirus phage YuA (A9J517, *E*-value 1e-72) as well as to ThyA of the host bacterium *R. solanacearum* (A3RUA7, *E*-value 6e-22). Despite the presence of a ThyA homolog, no homologs were found in ϕRSL1 for dihydrofolate reductase (EC 1.5.1.3), thymidylate kinase (EC 2.7.4.9), or ribonucleoside-diphosphate reductase α/β chains (EC 1.17.4.1), which are encoded by T4. Homologs to the T4 Alc, Ndd, endonuclease II (*denA*), and endonuclease IV (*denB*) proteins were not found in ϕRSL1.

NAD^+ salvage pathway

ϕRSL1 appears to encode pyridine nucleotide (NAD^+) salvage pathway. ϕRSL1 has a nicotinate phosphoribosyltransferase homolog (EC 2.4.2.11; ORF106) and an NMN adenyltransferase homolog (EC 2.7.7.1; ORF104), which likely serve for the synthesis of NAD^+ (Penfound and Foster, 1996). Being close to these ORFs, there is a homolog (ORF109) for Nudix hydrolases, which would recycle NADH back to precursors (Frick and Bessman., 1995), and a homolog (ORF101) for tripartite ATP-independent periplasmic transporter (TRAP-T) solute receptors. TRAP-T represents a novel type of secondary active transporter that functions in conjunction with an extracytoplasmic solute-binding receptor. The best characterized TRPT-T family member is that of *R. capsulatus* specific to C4-dicarboxylates (Forward et al., 1997). ORF101 may help transporting precursors of NAD^+. These ORFs appear to be assembled in an operon in region I and intermediate–late transcription was confirmed at least for ORF104 and ORF106. NAD^+-recruitment may help the activity of the ϕRSL1-encoded NAD^+-dependent DNA ligase (ORF065). A similar set of enzymes involved in the pyridine nucleotide salvage pathway has been reported for vibriophage KVP40 (Miller et al., 2003a).

Lysis proteins

ORF181 shows significant homology to chitinases (EC 3.2.1.14, family 19 glycoside hydrolase) from bacteria such as *Pseudomonas fluorescens* Pf-5 (Q4KA35, *E*-value 2e-27), *Burkholderia vietnamiensis* plasmid pBVIE05 (A4JWE6, *E*-value 4e-24), *Magnetospirillum gryphiswaldense* (A4TVX0, *E*-value 4e-21), and *Deinococcus radiodurans* (Q9RZ37, *E*-value of 2e-20). ORF181 also shows homology, albeit with lower statistical significance, with phage-encoded lytic enzymes such as R-type pyocin of *P. aeruginosa* (Q9S558, *E*-value 3e-18), lys of a prophage in *Xanthomonas campestris* pv. *campestris* (Q8P6J1, *E*-value 1e-16), and lys of a prophage in *S. maltophilia* K279a (B2FIY9, *E*-value 2e-16). We generated a sequence alignment of ORF181 with its close homologs and T4 lysozyme (EC 3.2.1.17) (Fig. 6). We found that ORF181 and its close homologs do not share the T4 lysozyme catalytic residues (E11 and D20) and stabilizing residues (Y18, T21, T26, E128, and T142) (Wozniak et al., 1994). On the other hand, the presence of the chitinase family 19 signature 2 ([LIVM]-[GSA]-F-x-[STAG](2)-[LIVMFY]-W-[FY]-W, PROSITE PS00774) was confirmed for ORF181

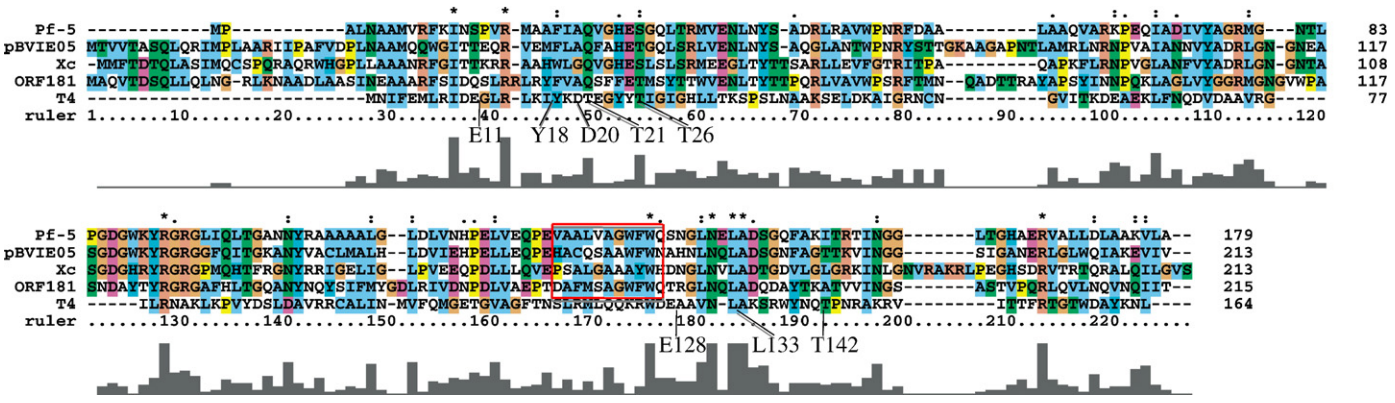

Fig. 6. Comparison of the amino acid sequence of ORF181 (chitinase-like protein) with related sequences in the databases. Pf-5, *P. fluorescens* Pf-5 (accession no. Q4KA35); pBVIE05, plasmid of *B. vietnamiensis* G4 (accession no. A4JWE6); Xc, *X. campestris* pv. *campestris* (accession no. Q8P6J1); T4, lysozyme of phage T4 (accession no. P00720). Catalytic and stabilizing residues of T4 lysozyme are indicated below the alignment. The chitinase family 19 signature 2 [LIVM]-[GSA]-F-x-[STAG](2)-[LIVMFY]-W-[FY]-W (accession no. PS00774) is indicated by a red box. Asterisks indicate identity and dots conservative substitutions. Conservative scores are also drawn below the alignment. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(Fig. 6). When ORF181 was cloned in *E. coli* BL21 with a plasmid vector pGEX4T-3 and its expression was induced by adding IPTG to the medium, the rod-shaped bacterial cells became round-shaped, frequently bound to each other and formed large aggregates (Supplementary Fig. S2). These results suggest that ϕRSL1 ORF181 encodes a chitinase-like bacterial lytic enzyme. In addition, ORF172 shows marginal amino acid sequence similarity to part of the α-1,3-glucosidase of *Bacillus circulans* (Q0WYG9, *E*-value 2e-04). The *B. circulans* enzyme contains carbohydrate binding modules of families 32 and 35 and a catalytic site of family 87 glycosyl hydrolases (Yano et al., 2006). It has been demonstrated that this enzyme, notably in combination with chitinase I, actively digests the cell wall and forms protoplasts from *Schizophyllum commune*.

Homing endonucleases

ORF122 and ORF332 of ϕRSL1 showed a conserved H-N-H motif of group I intron homing endonucleases from many bacteriophages (Lambowitz and Belfort, 1993; Clyman et al., 1994). The first prokaryotic case of a group I self-splicing intron was found in the thymidylate synthase gene of T4 phage (Chu et al., 1984). The T4 intron-encoded endonucleases are involved in the mobility (homing) of the intron into other sites. The presence of homing enzyme-like ORFs in ϕRSL1 suggests that this phage may contain group I intron or other related mobile sequences. We examined ORF230 encoding a thymidylate synthase-like sequence by comparing its DNA and amino acid sequences with those of other phages or bacteria. However, no sequence/structural features of group I intron or inteins were detected in this gene. In addition, ORF318 was similar to IS3/IS911 family insertion sequences.

Lipid metabolism

ϕRSL1 have several ORFs showing amino acid sequence similarities to proteins involved in lipid metabolism. Those include ORF083 (Acyl carrier protein, ACP, EC 1.3.99.3), and ORF261 (CDP-diacylglycerol-glycero-3-phosphate 3-phosphatidyltransferase, EC 2.7.8.5). CDP-diacylglycerol-glycero-3-phosphate 3-phosphatidyltransferase participates in glycerophospholipid metabolism and catalyzes the following reaction: CDP-diacylglycerol + sn-glycerol 3-phosphate = CMP + 3(3-sn-phosphatidyl)-sn-glycerol-sn-glycerol 1-phosphate. Currently, there is no information available about changes in lipid or lipopolysaccharide metabolism of phage-infected host cells or presence of lipid or lipoproteins in ϕRSL1 particles.

Carbohydrate metabolism and sulfonation

A cluster of ORFs in region IV (ORF232–ORF234–ORF235–ORF236–ORF237–ORF238) appears to be involved in carbohydrate metabolisms. ORF234 and ORF237 are similar to glucosyl transferase of bacteriophage ϕJL001 infecting a sponge-associated alpha-proteobacterium (Lohr et al., 2005). ORF235 is homologous to dolichol phosphate mannose (DPM) synthase (EC 2.4.1.83) of *Leifsonia xyli* subsp. *xyli* (Q6AH85, 1e-11). DPM synthase is required in N-glycosylation, O-mannosylation, and glycosylphosphatidylinositol membrane anchoring of protein. ORF236 is similar to the capsular polysaccharide biosynthesis protein of *Pasteurella piscicida* (Q8VW65, 6e-50). ORF238 is similar to adenylylsulfate kinases (EC 2.7.1.25). Adenylylsulfate kinase or ATP:adenosine-5′-phosphosulfate 3′-phosphotransferase is an enzyme involved in the sulfate activation pathway and performs the following reaction: ATP + adenylylsulfate = ADP + 3′-phosphoadenylylsulfate (PAPS) (Leyh et al., 1988). PAPS serves as the universal sulfonate donor molecule required for all sulfonation reactions (Farooqui, 1980). Many sugar sulfation reactions are also known to use PAPS as a co-factor (Strott, 2002). Interestingly, another ORF (ORF232) shows a significant sequence similarity to a carbohydrate sulfotransferase of *Xenopus tropicalis* (Q28G00, *E*-value 3e-5). Sulfated metabolites are known to serve as key modulators of interactions between plants and symbiotic/pathogenic bacteria (Schelle and Bertozzi, 2006), thus suggesting that these ϕRSL1 ORFs may contribute to the pathogenicity of its host.

Homospermidine synthase

Two ORFs encoded in region I appear to be involved in spermidine-related metabolism. ORF052 is similar to glutathionylspermidine synthase (EC 6.3.1.8) (A0YLT5, *E*-value 1e-108). This enzyme participates in glutathione metabolism and catalyzes the following reaction: glutathione spermidine + ATP = glutathionylspermidine + ADP + phosphate. ORF110 shows similarity to homospermidine synthases (EC 2.5.1.44) (A1SAY3, *E*-value of 1e-115). This enzyme is an NAD^+-dependent enzyme and catalyzes the formation of homospermidine from putrescine. Homospermidine, unlike the more common polyamines putrescine, spermidine and spermine, occurs in a few but distantly related organisms and is used as a specific chemotaxonomic marker in certain bacterial taxa, such as the alpha-proteobacteria including photosynthetic bacteria, *Agrobacterium*, *Rhizobium*, and cyanobacteria. In *R. solanacearum* (beta-proteobacterium), the occurrence of homospermidine is not known. The importance of polyamines including homospermidine in virus replication has been documented in literature (reviewed in Cohen, 1998). Virus-encoded enzymes involved in the polyamine synthesis were first found in a member of NCLDV, *Chlorella* virus PBCV-1 (Kaiser et al., 1999). Chlorella viruses are also known to have a giant dsDNA genome (300–380 kbp) encoding more than 300 genes, many of which are unexpected for a virus (Yamada et al., 2006). Although biological functions of virus-encoded polyamine synthesis enzymes and their products are currently unknown, it is interesting to note that a homospermidine synthase protein is a major component of PBCV-1 virion (Dunigan, D.D. et al., unpublished results), implying that the homospermidine synthase protein may have a structural role in the life cycle of PBCV-1.

MoxR AAA+ type chaperon system

ORF082 is similar to the RtcB protein of Halovirus HF1 (NP_861650, *E*-value 1e-31). RtcB is associated with RtcA protein that has an enzymatic activity of RNA 3′-terminal phosphate cyclase involved in the ATP-dependent conversion of the 3′-phosphate to the 2′, 3′-cyclic phosphodiester at the end of various RNA substrates (Genschik et al., 1998). ORF091 is similar to Mox-R-like ATPase of *Thermus* phage YS40 (NP_874121, *E*-value 2e-05). MoxR AAA+ ATPase family proteins generally function with Von Willebrand Factor Type A (VWA) domain-containing proteins to form a chaperone system, which has an important role in the folding or activation of proteins and protein complexes by mediating the insertion of metal cofactors into the substrate molecules. MoxR AAA+ family members are found throughout Bacteria and Archaea, but have not yet been detected in Eukaryota (Snider and Houry, 2006). Several ϕRSL1 ORFs are similar to proteins of the host bacteria or their relatives; ORF016 showed similarity to a transcription regulator (helix–turn–helix protein) of XRE family of *Ralstonia pickettii* (B2UJ28, *E*-value 2e-07). This ORF was expressed early in ϕRSL1 infection and most probably involved in the transcription regulation of ϕRSL1 genes. ORF171 is similar to the hemagglutinin-related autotransporter of *R. solanacearum* (Q8Y366, *E*-value 6e-21). ORF102 is similar to putative proteic killer active protein of *Limnobacter* sp. (*Burkholderiaceae*, accession no. A6GUP6, *E*-value 8e-05) or plasmid maintenance antidote protein of *Ralstonia eutropha* (Q0K0W4, *E*-value 0.03).

Phage-encoded tRNAs

The ϕRSL1 genome contains two tRNA genes [Gly (GGA) (positions 186,301–186,376, complement strand) and Pro (CCA) (positions 186,529–186,606, complement strand)] and one tRNA pseudogene showing a large sequence divergence from known tRNA sequences (HMM score = 2.98, see Materials and methods). In some bacteriophages, for example in T4 (Mosig, 1994), phage encoded tRNAs are

suggested to play a role in compensating codon-usage gaps between host and phage genomes. By encoding its own tRNA species corresponding to abundant phage codon types, phage may be able to efficiently replicate in a broad range of hosts. However, the codon usage of ϕRSL1 showed that both GGA-Gly (0.95%) and CCA-Pro (0.78%) are the rarest ones. These tRNAs might be used to incorporate unusual amino acids into phage proteins (Cathopoulis et al., 2007).

Transcription of ϕRSL1 genes

We performed an expression assay using a DNA microarray to characterize transcription patterns of ϕRSL1 genes during infection cycle. *R. solanacearum* MAFF730138 was used as a host, as this strain gave a better quality RNA preparation compared with strain M4S. Total RNA samples were extracted from MAFF730138 cells infected with ϕRSL1 at 10 min p.i., 30 min p.i., and 90 min p.i. We used a 40×55 array chip, with 1100 distinct sequence probes in duplicate. Each of the ORFs in the ϕRSL1 genome was represented by about 1 to 5 distinct probes of 35 to 40 nucleotide-long in the array chip. Only reproducible signals were considered in our analysis. Table 2 shows lists of ORFs with more than twofold increase or decrease in signal intensity across different times points. We found that most genes showing early expression (10–30 min p.i.) and later repression by 90 min p.i. (designated as early-intermediate genes) were confined within region I except for ORF170 in region II, ORF176, ORF188–ORF189–ORF191 in region III, and ORF236, ORF293–ORF294–ORF295 in region IV. In contrast, genes that showed increased expression during 30–90 min p.i. (designated as intermediate–late genes) were mostly concentrated around both extremities of region I, as well as in the entire region IV (Table 2 and Fig. 3). The intermediate–late genes also included several genes located in region II (ORF146, ORF147, ORF151, and ORF161) and region III (ORF171, ORF186 and ORF187). We also confirmed that putative genes for phage structural proteins were of intermediate–late genes. Genes involved in DNA metabolism were also intermediate–late, except for ORF065, which encodes NAD^+-dependent DNA ligase.

Prediction of transcriptional regulation and validation

ϕRSL1 ORFs are tightly spaced along the genome. However, there are several relatively large (>100 bp) and apparently noncoding regions in the genome. A spacer region between genes is usually occupied by regulatory sequences such as promoters and terminators. By taking advantage of expression data described above, we examined these long spacer regions for the presence of putative regulatory sequences. We identified several conserved sequences in upstream regions of early-intermediate genes. One typical example is shown in Fig. 7A, which represents one of the simplest cases among those we identified. A 100-bp region between the translation stop codon of ORF036 and the initiation codon of the following ORF037 (positions 110,482–110,760, region I) contains a potential stem-loop terminator for ORF036 and three sequence elements, GCATAAT, GTACAC, and CTCCCATGCGAAACGGAG that are conserved across different spacers. The sequence of GGAG of the last element overlaps the SD of ORF037. Exactly the same elements placed in the same order were found in the intergenic regions of ORF017–ORF018, ORF059–ORF061, and ORF299–ORF300, all of which are located in the region I encoding many early-intermediate genes. Similar elements were also found in the upstream regions of ORF115 and ORF295 as well as in the intergenic regions of ORF086–ORF087, ORF187–ORF188, ORF294–ORF293, ORF315–ORF316, and ORF324–ORF325. ORFs downstream of these putative regulatory sequences were found to be expressed as early-intermediate genes. In contrast, these sequence elements were not always found in the upstream region of early-intermediate genes. ORF232 and ORF295 (region IV) seem to contain usual σ^{70}-type promoter elements (data not shown). ORF295 thus may contain multiple promoter elements.

As for the intermediate–late gene expression, we identified two sets of possible promoter elements. The first one is represented by that found in an intergenic region between ORF221 and ORF219 located in region IV. As seen in Fig. 7B, within an approximately 90-bp region, there is a ρ-independent terminator-like stem-loop element, and a CTTCAT element approximately 20-bp upstream of SD sequence of

Table 2
Expression of ϕRSL1 genes determined by DNA microarray analysis.

ORFs showing more than twofold increase in expression				ORFs showing more than twofold decrease in expression			
10 min p.l.–90 min p.l.		30 min p.l.–90 min p.l.		10 min p.l.–90 min p.l.		30 min p.l.–90 min p.l.	
ORF008	ORF204	ORF058	ORF226	ORF018	ORF305	ORF018	ORF317
ORF015	ORF209	ORF104	ORF230	ORF042	ORF306	ORF019	ORF318
ORF035	ORF219	ORF106	ORF232	ORF045	ORF307	ORF020	ORF319
ORF058	ORF222	ORF107	ORF246	ORF053	ORF315	ORF024	ORF320
ORF059	ORF230	ORF128	ORF249	ORF055	ORF316	ORF025	ORF338
ORF070	ORF249	ORF129	ORF250	ORF061		ORF026	ORF340
ORF104	ORF250	ORF132	ORF253	ORF065		ORF028	
ORF106	ORF253	ORF133	ORF256	ORF073		ORF034	
ORF107	ORF256	ORF134	ORF258	ORF087		ORF037	
ORF128	ORF258	ORF135	ORF259	ORF091		ORF039	
ORF132	ORF265	ORF136	ORF265	ORF096		ORF041	
ORF133	ORF266	ORF137	ORF266	ORF115		ORF042	
ORF134	ORF270	ORF140	ORF271	ORF116		ORF052	
ORF135	ORF271	ORF141	ORF280	ORF117		ORF053	
ORF136	ORF280	ORF142	ORF281	ORF119		ORF063	
ORF137	ORF281	ORF146	ORF282	ORF170		ORF065	
ORF140	ORF282	ORF147	ORF287	ORF173		ORF087	
ORF141	ORF289	ORF151	ORF299	ORF188		ORF089	
ORF142	ORF299	ORF171	ORF326	ORF189		ORF091	
ORF146	ORF326	ORF187	ORF327	ORF191		ORF293	
ORF147	ORF327	ORF203	ORF329	ORF236		ORF300	
ORF151	ORF329	ORF204	ORF335	ORF294		ORF301	
ORF161	ORF331	ORF209		ORF295		ORF303	
ORF171	ORF334	ORF213		ORF297		ORF305	
ORF186	ORF341	ORF219		ORF300		ORF307	
ORF187	ORF342	ORF221		ORF301		ORF315	
ORF203		ORF222		ORF303		ORF316	

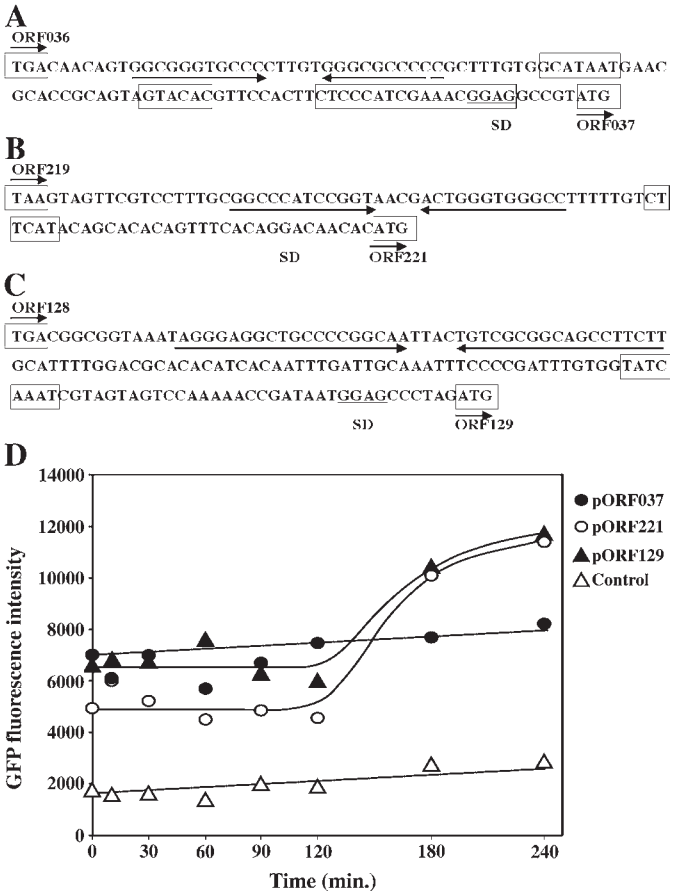

Fig. 7. Promoter sequences of ϕRSL1 genes. (A) The ORF036–ORF037 region containing a ρ-independent terminator (underlined) and three promoter elements (boxed). The termination codon (TGA) of ORF36, the initiation codon of ORF37, and Shine Dalgarno sequence are indicated. ORF37 showed an early-intermediate gene expression pattern. (B) The ORF221–ORF221 region (on the complementary strand) is shown. Elements are indicated as in (A). ORF221 showed an intermediate–late expression pattern. (C) The ORF128–ORF129 region is shown. Elements are indicated as in (A). ORF129 showed an intermediate–late expression pattern. (D) Expression of GFP by ϕRSL1 promoters. *R. solanacearum* cells transformed by pRSS12–ORF037 promoter (pORF037), pRSS12–ORF221 promoter (pORF221), or pRSS12–ORF129 promoter (pORF129) were infected with ϕRSL1 and GFP fluorescence emission from the cells was monitored at several times post infection. Control, non-transformed *R. solanacearum* cells.

ORF219. A similar set of elements was also found in the intergenic region between ORF223–ORF222, but in this case, the possible promoter sequence was CATCAT. Both ORF221 (in an operon with ORF219) and ORF223 showed intermediate–late expression patterns (Table 2). This type of elements (a terminator-CTTCAT located 20–30 bp upstream of SD sequence) were detected in the intergenic regions of ORF034–ORF035 (TCACAT), ORF209–ORF208 (TTCCAT), ORF228–ORF226 (CTTCAT), ORF231–ORF230 (TTCCAT), ORF233–ORF232 (CTTCATCAT), ORF261–ORF259 (CCTCAT), ORF273–ORF271 (CATCAT), and ORF340–ORF341 (CATCAT). All of the ORFs following these putative regulatory elements (such as ORF035) and their following ORFs in an operon showed an intermediate–late expression patterns (Table 2). It is noteworthy that most ORFs with these elements are located in region IV. The other set of elements is illustrated by the one found in the intergenic region of ORF128–ORF129 (Fig. 7C). A sequence element TATCAAAT is present downstream of a ρ–independent terminator-like stem-loop and followed by 25 nucleotides before SD-sequence. Similar elements (a terminator-TATCAAAT-like sequence located 20–30 bp upstream of SD sequence) were also detected in intergenic regions of ORF057–ORF058 (ATACAAAG), ORF69–ORF70 (TACAAAA), ORF139–ORF140 (CACAAAC), ORF170–ORF171 (CATCAAAT), ORF298–ORF299 (ATCCAAAT), ORF324–ORF325 (TTCAGAAT), ORF328–ORF329 (CATCAAGG), and ORF334–ORF335 (TCGCAAAC). All the ORFs immediately downstream of these elements and their following ORFs in an operon showed intermediate–late expression. Most of these genes are confined in region I of the ϕRSL1 genome.

We examined the activity of the predicted three types of promoters using a GFP-expressing single-copy plasmid pRSS12 (Kawasaki et al., 2007, 2009), where the *lac* promoter for *gfp* expression was replaced with a ϕRSL1 promoter sequence. Bacterial cells transformed with plasmids pORF037 (containing the predicted early-intermediate promoter of ORF037), pORF221 (containing the first type of the predicted intermediate–late promoter of ORF221), and pORF129 (containing the second type of the predicted intermediate–late promoter of ORF129) showed GFP fluorescence at a relative intensity (with a control of *lac* as 100) of 350, 310, and 250, respectively. After infection with ϕRSL1, GFP fluorescence intensity retained almost the same levels in cells containing pORF037, whereas cells with pORF221 or pORF129 showed increased GFP fluorescence after 120 min p.i. (a 30- to 60-min time lag from DNA microarray data) (Fig. 7D). Increased expression of GFP driven by the promoter of ORF221 or ORF129 at late stages of ϕRSL1 infection was also observed for plaques formed on plates, where only cells at the rim of each plaque showed strong GFP fluorescence in both cases (Supplementary Fig. S3). In the case of pORF037-transformed cells, such a halo was not visible but the bacterial lawn itself showed GFP fluorescence. These results indicate that the promoter sequence of the early-intermediate gene ORF037 functions to express GFP even without ϕRSL1 infection, and that the promoter sequences of the intermediate–late genes ORF221 and ORF129 actually enhanced the expression of GFP after infection with ϕRSL1 in accordance with the microarray data.

Genome evolution

In addition to homologs for the core genes of T4-like phages (Filee et al., 2006), the ϕRSL1 genome encodes over 300 different genes. Our phylogenetic analysis result of PolA, the lack of sequence similarity between the major capsid protein (ORF139) and known capsid proteins, and the remarkable abundance of ORFans together suggest that ϕRSL1 represents a new lineage of Myoviridae. Although functions could not be predicted for most of those genes due to the lack of homologs in the current databases, we could assign predicted functions for some of them. Among them, putative genes for enzymes involved in the lipid/carbohydrate metabolisms and the pyridine nucleotide salvage pathway are remarkable. These and other genes might provide the phage with selective advantages and possibly play important roles in its stable lytic infection capability, a characteristic feature of ϕRSL1. Some ϕRSL1-encoded proteins show detectable homology to a broad phylogenetic range of microorganisms, as well as in different groups of phages and viruses (Fig. 4B and Table 2), suggesting a very ancient origin of ϕRSL1. Non-core genes are found interspersed among clusters of core genes across the entire genome (Fig. 4). For jumbo phages like ϕRSL1, vastly large parts of their genomes are occupied by non-core genes. According to the headful DNA packaging mechanism of T4 (Black et al., 1994), the capsid size determines the size of the genome to be densely packaged in it; namely an increase in capsid size is required to pack a larger genome. Consistently, the dimension of ϕRSL1 capsid (ϕ150 nm) is significantly larger than that of T4 (90 nm × 112 nm). To increase genome size, a phage may need innovational changes in the capsid proteins. We identified six proteins including two major proteins gp136 (33.0 kDa) and gp135 (24.7 kDa) in the ϕRSL1 virion, which did not show significant sequence similarities to previously identified phage capsid proteins. Determining the interaction and assembling of these proteins in the virion and comparison of their architecture to other phage capsid structures would be a very interesting future issue.

The ϕRSL1 genome consists of four major genomic regions, each of which harbors many ORFs in the same orientation (Fig. 3). T4-like phages shows two genomic regions encoding genes for DNA replication and for virion structural proteins, respectively (Filee et al., 2006). In ϕRSL1, the corresponding core genes are distributed in region II and the left end of region IV for DNA replication proteins and the right end of region I and the left end of region IV for structural proteins. This gene organization suggests an ancestral form of ϕRSL1 genome, in which the region II was connected to the right end of region IV and the right end of region I to the left end of region IV. This hypothesis is consistent with the distribution of specific promoter elements as described above. Currently, only a handful of jumbo phages have been isolated. Jumbo phages might have been missed in ordinary screenings due to their large size because they diffuse too slowly in the top agar gels typically used to plaque phages. Reducing the top agar concentration from 0.7–0.8% to 0.45–0.5% allowed us to plaque ϕRSL1. A similar observation was reported by Serwer et al. (2007) for *Bacillus* phage G. It is very likely that more jumbo phages will be isolated by adapting these plaquing conditions.

Conclusion

The genome of ϕRSL1 (231,255 bp encoding 343 ORFs) is vastly different from previously studied dsDNA phages, and our results suggest that this phage represents an evolutionarily distinct branch of the Myoviridae. The majority of proteins predicted from the ϕRSL1 genome sequence have no obvious homologs in the current sequence databases but some showed significant similarity to enzymes involved in lipid and carbohydrate metabolisms, which are rare in phages. The genome information of ϕRSL1 will serve as a ground to understand many aspects of its structure, biochemistry, physiology and evolution. Especially the molecular basis of its stable and long-lasting lytic infection can be understood. The complete genome sequence of this phage readily permits the cloning of individual genes, and help the construction of mutants and understanding of the functions of many of its genes.

Materials and methods

Bacterial strains and phages

R. solanacearum strains M4S (Tanaka et al., 1990), MAFF301558, and MAFF730138 (Yamada et al., 2007) were used as hosts for ϕRSL1 propagation. The bacterial cells were cultured in CPG medium containing 0.1% casamino acids, 1% peptone, and 0.5% glucose (Horita and Tsuchiya, 2002) at 28 °C with shaking at 200–300 rpm. An overnight culture of bacterial cells grown in CPG medium was diluted 100-fold with 100 ml fresh CPG medium in a 500 ml flask. To collect sufficient phage particles, a total of 2 L of bacterial culture was grown. When the cultures reached 0.2 unit of OD_{600}, the phage was added at a multiplicity of infection (moi) of 0.01–1.0. After further growth for 9–18 h, the cells were removed by centrifugation with a R12A2 rotor in a Hitachi himac CR21E centrifuge (Hitachi Koki Co. Ltd., Tokyo, Japan), at 8000×*g* for 15 min at 4 °C. The supernatant was passed through a 0.45 μm membrane filter and phage particles were precipitated by centrifugation with a P28S rotor in a Hitachi CΠ100β centrifuge at 40,000×*g* for 1 h at 4 °C and dissolved in SM buffer (50 mM Tris–HCl at pH 7.5, 100 mM NaCl, 10 mM $MgSO_4$, and 0.01% gelatine). Purified phages were stored at 4 °C until use. Bacteriophage particles purified by CsCl-gradient ultracentrifugation (with a P28S rotor in a Hitachi CP100β ultracentrifuge) were stained with Na-phosphotungstate before observation in a Hitachi H600A electron microscope according to Dykstra (1993). λ phage particles were used as an internal standard marker for size determination. *E. coli* XL10-Gold and pBluescript II SK (+) were obtained from Stratagene (La Jolla, CA). *E. coli* BL21 and pGEX4T-3 were from Amersham Biosciences (Uppsala, Sweden).

Single-step growth experiment

Single-step growth experiments were performed as described (Carlson, 1994; Kawasaki et al., 2009) with some modifications. Cells of *R. solanacearum* M4S (0.5 U of OD_{600}) were harvested by centrifugation and resuspended in the fresh CPG (ca. 3×10^8 CFU/ml). Phage ϕRSL1 was added at a moi of 0.01 and allowed to adsorb for 5 min at room temperature. After centrifugation and resuspending in CPG, the cells were incubated at 28 °C. Samples were taken at 30 min-intervals up to 6 h and immediately diluted with or without chlorofolm treatment, and the titers were determined by the double-layered agar plate method.

DNA manipulations and sequencing

Standard molecular biological techniques for DNA isolation, digestion with restriction enzymes and other nucleases, and construction of recombinant DNAs were followed according to Sambrook and Russell (2001). Phage DNA was isolated from the purified phage particles by phenol extraction. For genome size determination, the purified phage particles were embedded in 0.7% low-melting-point agarose (InCert agarose, FMC Corp, Philadelphia, PA) and after treatment with proteinase K (1 mg/ml, Merck Ltd., Tokyo, Japan) and 1% Sarkosyl, subjected to pulsed-fields gel electrophoresis with a CHEF MAPPER electrophoresis apparatus (Bio-Rad Lab., Hercules, CA) according to Higashiyama and Yamada (1991). Shotgun cloning and sequencing was performed at Hitachi High-Tech Fields Corp. (Tsukuba, Japan) as follows: ϕRSL1 whole genomic DNA was fragmented by sonication. DNA fragments of ~2 kb range were blunt-ended and cloned with pTS1/*Hin*cII vector (NipponGene, Wako, Osaka, Japan) in *E. coli* cells. Shotgun sequencing of the clones (with an averaged insert of 2.0 kb) was performed with a BigDye Terminator ver 3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA) in Applied Biosystem 3700 DNA Analyzer. A total of 3648 sequences larger than 150 bases were assembled by the use of a phred/phrap/consed program (http://www.phrap.org). These sequences (2,718,115 bp) were assembled into 9 contigs (contig 1 of 56,736 bp in size; contig 2 of 46,325 bp; contig 3 of 26,980 bp; contig 4 of 26,855 bp; contig 5 of 21,477 bp; contig 6 of 16,728 bp; contig 7 of 11,531 bp; contig 8 of 9,039 bp; contig 9 of 7,594 bp). Bridge clones obtained for contig 2-contig 5, contig 3-contig 4, and contig 7-contig 9 were used to fill the gaps by primer walking. The remaining gaps were also filled by primer walking using pure ϕRSL1 DNA as template. With combinations of primers corresponding to each contig end, all possibilities were tried by PCR. For each gap, at least two different sets of primers were used to ensure the connection. Twenty-five rounds of PCR were performed with 1 ng of ϕRSL1 DNA as a template under the standard conditions in a MY Cycler (Bio-Rad). Nucleotide sequence of these DNA fragments was determined as described above. The analyzed sequences corresponded to 12 times the final genome size of 231,255 bp. Potential ORFs larger than 300 bp were identified using Glimmer (Delcher et al., 1999) and GeneMark (Besemer et al., 2001). Homology searches were performed using BLAST/RPS-BLAST (Altschul et al., 1997) against the UniProt sequence database (UniProt Consortium, 2007) and the NCBI/CDD database (Wheeler et al., 2007). The tRNA genes were identified using tRNAscan-SE (Lowe and Eddy, 1997). Multiple-sequence alignments were generated using the DNASIS program (version 3.6; Hitachi Softwear Engineering, Co., Ltd., Tokyo, Japan). A circular genome map was drawn using CGView (Stothard and Wishart, 2005).

For cloning and expression of ORF184 encoding putative chitinase in *E. coli* cells, a 575 bp fragment was amplified from ϕRSL1 DNA by PCR with a forward primer, 5′-A*GG ATC* CAT GCG GCT CCA TCG GTC CAA GGA and a reverse primer, 5′-CGA TAG GA*G CGG CCG* CTC TAT TCC TGC TTG. The PCR products were treated with T4 polynucleotide kinase, ligated to the *Hin*cII site of pTWV228 (a low copy number

plasmid, Takara Bio Inc., Tokyo, Japan)) and introduced into *E. coli* XL10-Gold cells. pTWV228–ORF181 isolated from a transformant was digested with *Bam*HI and *Not*I and an approximately 600 bp fragment was ligated to *Bam*HI/*Not*I sites of pGEX4T-3. Exact in-frame connection of ORF181 in pGEX4T-3 was confirmed by direct sequencing of the plasmid. The plasmid was introduced into *E. coli* BL21 for expression analyses.

DNA microarray analysis

Total RNA was isolated from 3 ml of ϕRSL1-infected MAFF730138 cells (containing 1×10^8 CFU infected at moi 4) at 10, 30, and 90 min post infection (p.i.) using an RNAprotect Bacteria Reagent kit (Qiagen K.K., Tokyo, Japan) according to the manufacturer's protocol. The integrity and concentration of total RNA were measured using a bioanalysis unit (Agilent 2100 Bioanalyser, Aligent Tech. Palo Alto, CA). Fluorescence-labeled antisense RNA was synthesized by direct incorporation of Cy5-UTP (GE Healthcare Bio-Science Corp., Piscataway, NJ), using each RNA sample and an RNA Transcript SureLABEL Core kit (Takara Bio Inc.). The labeled antisense RNAs were hybridized simultaneously to the microarray chip. DNA microarray preparation, hybridization, processing, scanning, and analyses were performed by Filgen Inc. (Nagoya Japan). For each ϕRSL1 ORF, 1 to 5 coding sequences of 35 to 40 nucleotide-long were synthesized. A total of 1,100 sequences were put in duplicate on a 40×55 array chip. The fluorescence images of hybridized microarrays were obtained with a GenePix 4000B scanner (Molecular Devices, Sunnyvale, CA). The Array-Pro Analyzer Ver4.5 (Media Cybernetics, Inc., Silver Spring, MD) was used to determine the signal intensity of each spot and its local background. Scan data images were analyzed using Microarray Data Analysis Tool Ver3.0 software (Filgen Inc.)

Southern and dot blot hybridization

Genomic DNA from *R. solanacearum* cells was prepared by minipreparation method according to Ausubel et al. (1995). After digestion with various restriction enzymes, DNA fragments were separated by agarose gel electrophoresis, blotted onto a nylon membrane (Biodyne, Pall Gelman Laboratory, Closter, NJ), hybridized with a probe (ϕRSL1 genomic DNA) labeled with fluorescein (Gene Imeges Random Prime labeling kit; Amersham Biosciences Uppsala, Sweden), and detected with a Gene Images CDP-Star detection module (Amersham Biosciences). For dot blot hybridization to detect nascent phage DNA synthesis in infected cells, the total DNA (1.0 μg) was directly blotted onto a nylon membrane. Hybridization was performed in buffer containing $5\times$SSC, 0.1% SDS, 5% liquid block and 5% dextran sulfate for 16 h at 65 °C. The filter was washed at 60 °C in $1\times$SSC and 0.1% sodium dodecylsulfate (SDS) for 15 min, and then in 0.5% SSC and 0.1% SDS for 15 min with agitation according to the manufacturer's protocol. The hybridization signals were detected by exposing the filter onto an X-ray film (RX-U, Fuji Film, Tokyo, Japan).

SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and peptide sequencing

Purified phage particles were subjected to SDS-polyacrylamide gel electrophoresis (PAGE) (12% polyacrylamide) according to Laemmli (1970). The separated proteins were transferred to polyvinylidene difluoride (PVDF) nylon membranes (Immobilon, Nihon Millipore K.K., Tokyo, Japan) with a semi-dry transfer cell (Bio-Rad). Well separated protein bands were subjected to N-terminal peptide sequence analysis on a protein sequencer (Shimadzu Biotech, Kyoto, Japan). First 5 amino acid residues were read for each protein sample.

Promoter activity assay

Putative promoter sequences detected on the ϕRSL1 genomic DNA were amplified by PCR with appropriate primer sets (ORF037F, 5′-GCG GAA TTC CAA CAG TGG CGG GTG C and ORF037R, 5′-TTG GAA TTC CGG CGG TAA ATA GGG AGG; ORF221F, 5′-GTA GAA TTC CCT TTG CGG CCC ATC C and ORF221R, 5′-GTG TTG TCC TGT GAA ACT GTG TGC TGT AT; ORF129F, 5′-TTG GAA TTC CGG CGG AAA TAG GGA GG and ORF129R, 5′-TAG GGC TCC ATT ATC GGT TTT TGG ACT A). The PCR products including the promoter sequences for ORF037 (positions 16815–16914), ORF221 (positions 157796–157721), and ORF129 (positions 70779–70923), were individually connected to the *Eco*RI site of pRSS12 according to Kawasaki et al. (2007 and 2009), replacing the *lac* promoter to express the *gfp* reporter gene. After introducing the plasmid into cells of *R. solanacearum* cells by electroporation as described previously (Kawasaki et al., 2007), transformants were selected on CPG plates containing 50 μg/ml kanamycin (Meiji Seika, Tokyo, Japan). GFP fluorescence of transformed cells (3×10^7 cells/ml) was measured using an FP-6500 spectrofluorometer (Jasco, Essex, UK) or a microplate reader Infinite 200, Wako, Osaka Japan) at 395 nm excitation and 509 nm emission wave lengths. In some experiments, the GFP fluorescence intensity was monitored at appropriate intervals during infection by ϕRSL1.

Bioinformatics analysis

Protein coding genes were predicted using and Glimmer (Delcher et al., 1999). Homology searches were performed using BLAST/RPS-BLAST (Altschul et al., 1997) against the UniProt sequence database (Boutet et al., 2007), the NCBI Genbank/env and the NCBI/CDD database (Marchler-Bauer et al., 2009). tRNA genes were identified using tRNAscan-SE (Lowe and Eddy, 1997). tRNAscan-SE reports three scores: Cove score (based on sequence and secondary structure conservation), HMM score (based on sequence conservation) and secondary structure score. A tRNA-like sequence with an HMM score below 10 bits or with a secondary structure score below 5 bits was considered as a pseudogene as implemented in tRNAscan-SE. A circular genome map was drawn using CGView (Stothard and Wishart, 2005). ORFans were defined as predicted genes lacking detectable homologs in the UniProt or CDD databases with a threshold *E*-value of 0.001.

Nucleotide sequence accession number

The sequence data for ϕRSL1 have been deposited in the DDBJ database under accession no. AB366653 (updated on 9 June, 2009).

Acknowledgments

We thank Jean-Michel Claverie (head of Information Genomique et Structurale) for laboratory space and support for the computer analyses. This study was supported in part by the Industrial Technology Research Grant Program in 04A09505 from the New Energy and Industrial Technology Development Organization (NEDO) of Japan and by a Grant-in Aid from the Ministry of Education, Science, Sport and Culture of Japan (21580095 to T. Y.).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.virol.2009.11.043.

References

Ackermann, H.W., 2003. Bacteriophage observations and evolution. Res. Microbiol. 154, 245–251.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.

Ashelford, K.E., Day, M.J., Fry, J.C., 2003. Elevated abundance of bacteriophage infecting bacteria in soil. Appl. Environ. Microbiol. 69, 285–289.

Ausubel, F., Brent, R., Kjngston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., Struhl, K., 1995. Short protocols in molecular biology, 3rd ed. John Wiley and Sons, Inc., Hoboken, NJ.

Besemer, J., Lomsadze, A., Borodovsky, M., 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory region. Nucleic Acids Res. 29, 2607–2618.

Black, L.W., Showe, M.K., Steven, A.C., 1994. Morphogenesis of the T4 head. In: Karam, J., Drake, J.W., Kreuzer, K.N., Mosig, G., Hall, D.H., Eiserlig, F.A., Black, L.W., Spicer, E.K., Kutter, E., Carlson, K., Miller, E.S. (Eds.), Molecular biology of bacteriophage T4. ASM Press, Washington, D. C., pp. 218–258.

Boutet, E., Lieberherr, D., Tognolli, M., Schneider, M., Bairoch, A., 2007. UniProt KB/Swiss-Prot. Methods Mol. Biol. 406, 89–112.

Carlson, K., 1994. Single-step growth. In: Karam, J., Drake, J.W., Kreuzer, K.N., Mosig, G., Hall, D.H., Eiserlig, F.A., Black, L.W., Spicer, E.K., Kutter, E., Carlson, K., Miller, E.S. (Eds.), Molecular biology of bacteriophage T4. ASM press, Washington, D. C., pp. 434–437.

Cathopoulis, T., Chuawong, P., Hendrickson, T.L., 2007. Novel tRNA aminoacylation mechanisms. Mol. Biosyst. 3, 408–418.

Chang, H.-C., Chen, C.-R., Lin, J.-W., Shen, G.-H., Chang, K.-M., Tseng, Y.-H., Weng, S.-F., 2005. Isolation and characterization of novel giant *Stenotrophomonas maltophila* phage ϕSMA5. Appl. Environ. Microbiol. 71, 1387–1393.

Chu, F.K., Maley, G.F., Maley, F., Belfort, M., 1984. Intervening sequence in the thymidylate synthase gene of bacteriophage T4. Proc. Natl. Acad. Sci. U.S.A. 81, 3049–3053.

Clyman, J., Quirk, S., Belfort, M., 1994. Mobile introns in the T-even phages. In: Karam, J., Drake, J.W., Kreuzer, K.N., Mosig, G., Hall, D.H., Eiserlig, F.A., Black, L.W., Spicer, E.K., Kutter, E., Carlson, K., Miller, E.S. (Eds.), Molecular biology of bacteriophage T4. ASM press, Washington, D. C., pp. 83–88.

Cohen, S.S., 1998. A guide to the polyamines. Oxford University Press, New York, p. 595.

Delcher, A.L., Harmon, D., Kasif, S., White, O., Salzberg, S.L., 1999. Improved microbial gene identification with GLIMMER. Nucleic Acids Res. 27, 4636–4641.

Dykstra, M.J., 1993. A manual of applied technique for biological electron microscopy. Plenum Press, New York.

Farooqui, A.A., 1980. 3′-Phosphoadenosine 5′-phosphosulfate: metabolism in mammalian tissues. Int. J. Biochem. 12, 529–535.

Filee, J., Bapteste, E., Susko, E., Krisch, H.M., 2006. A selective barrier to horizontal gene transfer in the T4-type bacteriophages that has preserved a core genome with the viral replication and structural genes. Mol. Biol. Evol. 23, 1688–1696.

Forward, J.A., Behrendt, M.C., Wyborn, N.R., Cross, R., Kelly, D.J., 1997. TRAP transporters: a new family of periplasmic solute transport system encoded by the dctPQM genes of *Rhodobacter capsulatus* and by homologs in diverse gram-negative bacteria. J. Bacteriol. 179, 5482–5493.

Frick, D.N., Bessman, M.J., 1995. Cloning, purification, and properties of a novel NADH pyrophosphatase. J. Biol. Chem. 270, 1529–1534.

Gabriel, D.W., Allen, C., Schell, M., Denny, T.P., Greenberg, J.T., Duan, Y.P., Flores-Cruz, Z., Huang, Q., Clifford, J.M., Presting, G., Gonzalez, E.T., Reddy, J., Elphinstone, J., Swanson, J., Yao, J., Mulholland, V., Liu, L., Farmerie, W., Patnaikuni, M., Balogh, B., Norman, D., Alvarez, A., Castillo, J.A., Jones, J., Saddler, G., Walunas, T., Zhukov, A.A., Mikhailova, N., 2006. Identification of open reading frames unique to a select agent: *Ralstonia solanacearum* race 3 biovar 2. Mol. Plant-Microbe Interact. 19, 69–79.

Genschik, P., Drabikowski, K., Filipowicz, 1998. Characterization of the *Escherichia coli* RNA-3′-terminal phosphate cyclase and its σ^{54}-regulated operon. J. Biol. Chem. 273, 25516–25526.

Greenberg, R.G., He, P., Hilfinger, J., Tseng, M.-J., 1994. Deoxynucleoside triphosphate synthesis and phage T4 DNA replication. In: Karam, J., Drake, J.W., Kreuzer, K.N., Mosig, G., Hall, D.H., Eiserlig, F.A., Black, L.W., Spicer, E.K., Kutter, E., Carlson, K., Miller, E.S. (Eds.), Molecular biology of bacteriophage T4. ASM press, Washington, D.C., pp. 14–27.

Hayward, A.C., 2000. *Ralstonia solanacearum*. In: Lederberg, J. (Ed.), Encyclopedia of Microbiology, vol. 4. Academic Press, San Diego, C. A., pp. 32–42.

Hendrix, R.W., 1999. Evolution: the long evolutionary reach of viruses. Curr. Biol. 9, R914–R917.

Hendrix, R.W., 2002. Bacteriophages: evolution of the majority. Theor. Popul. Biol. 61, 471–480.

Hendrix, R.W., 2009. Jumbo bacteriophages. Curr. Top. Microbiol. Immunol. 328, 229–240.

Hertveldt, K., Lavigne, R., Pleteneva, E., Sernova, N., Kurochkina, L., Korchevskii, R., Robben, J., Mesyanxhinov, V., Krylov, V.N., Volckaert, G., 2005. Genome comparison of *Pseudomonas aeruginosa* large phages. J. Mol. Biol. 354, 536–545.

Higashiyama, T., Yamada, T., 1991. Electrophoretic karyotyping and chromosomal gene mapping of *Chlorella*. Nucleic Acids Res. 19, 6191–6195.

Horita, M., Tsuchiya, K., 2002. Causal agent of bacterial wilt disease *Ralstonia solanacearum*. In: National Institute of Agricultural Sciences (Ed.), MAFF microorganism genetic resources mannual No. 12. National Institute of Agricultural Sciences, Tsukuba, Japan, pp. 5–8.

Kaiser, A., Vollmet, M., Tholl, D., Graves, M.V., Gurnon, J.R., Xing, W., Lisec, A.D., Nickerson, K.W., Van Etten, J.L., 1999. Chlorella virus PBCV-1 encodes a functional homospermidine synthase. Virology 263, 254–262.

Kawasaki, T., Satsuma, H., Fujie, M., Usami, S., Yamada, T., 2007. Monitoring of phytopathogenic *Ralstonia solanacearum* cells using green fluorescent protein-expressing plasmid derived from bacteriophage ϕRSS1. J. Biosci. Bioeng. 104, 451–456.

Kawasaki, T., Shimizu, M., Satsuma, H., Fujiwara, A., Fujie, M., Usami, S., Yamada, T., 2009. Genomic characterization of *Ralstonia solanacearum* phage ϕRSB1, a T7-like wide-host-range phage. J. Bacteriol. 191, 422–427.

Kiljunen, S., Hakala, K., Pinta, E., Huttunen, S., Pluta, P., Gador, A., Lonnberg, H., Skurnik, M., 2005. Yersinophage ϕR1-37 is a tailed bacteriophage having a 270 kb DNA genome with thymidine replaced by deoxyuridine. Microbiology 151, 4093–4102.

Laemmli, U.K., 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227, 680–685.

Lambowitz, A.M., Belfort, M., 1993. Introns as mobile genetic elements. Annu. Rev. Biochem. 622, 587–622.

Leyh, T.S., Taylor, J.C., Markham, G.D.., 1988. The sulfate activation locus of *Escherichia coli* K12: cloning, genetics, and enzymatic characterization. J. Biol. Chem. 263, 2408–2416.

Lohr, J.E., Chen, F., Hill, R.T., 2005. Genomic analysis of bacteriophage ϕJL001: insight Into its interaction with a sponge-associated alpha-proteobacterium. Appl. Environ, Microbiol. 71, 1598–1609.

Lowe, T.M., Eddy, S.R., 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25, 955–964.

Marchler-Bauer, A., Anderson, J.B., Chitsaz, F., Derbyshire, M.K., DeWeese-Scott, C., Fong, J.H., Geer, L.Y., Geer, R.C., Gonzales, N.R., Gwadz, M., He, S., Hurwitz, D.I., Jackson, J.D., Ke, Z., Lanczycki, C.J., Liebert, C.A., Liu, C., Lu, F., Lu, S., Marchler, G.H., Mullokandov, M., Song, J.S., Tasneem, A., Thanki, N., Yamashita, R.A., Zhang, D., Zhang, N., Bryant, S.H., 2009. CDD: specific functional annotation with the Conserved Domain Database. Nucleic Acids Res. 37, D205–210.

Martin, M.O., Long, S.R., 1984. Generalized transduction of *Rhizobium meliloti*. J. Bacteriol. 159, 125–129.

Mesyanzhinov, V.V., Robben, J., Grymonprez, B., Kostyuchenko, V.A., Bourkaltseva, M.V., Sykilinda, N.N., Krylov, V.N., Volckaert, G., 2002. J. Mol. Biol. 317, 1–19.

Miller, E.S., Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Durkin, A.S., Ciecko, A., Feldblyum, T.V., White, O., Paulsen, I.T., Nierman, W.C., Lee, J., Szczypinski, B., Fraser, C.M., 2003a. Complete genome sequence of the broad-host-range vibriophage KVP40: comparative genomics of a T4-related bacteriophage. J. Bacteriol. 185, 5220–5233.

Miller, E.S., Kutter, E., Mosig, G., Arisaka, F., Kunisawa, T., Ruger, W., 2003b. Bacteriophage T4 genome. Microbiol. Mol. Biol. Rev. 67, 66–85.

Mosig, G., 1994. Synthesis and maturation of T4-encoded tRNAs. In: Karam, J., Drake, J.W., Kreuzer, K.N., Mosig, G., Hall, D.H., Eiserlig, F.A., Black, L.W., Spicer, E.K., Kutter, E., Carlson, K., Miller, E.S. (Eds.), Molecular biology of bacteriophage T4. ASM press, Washington, D.C., pp. 182–185.

Penfound, T., Foster, J.W., 1996. Biosynthesis and recycling of NAD. In: Niedhardt, F.C., Curtiss, R., Ingraham, J.L., Lin, E.C.C., Low, K.B., Magasanik, B., Reznikoff, W.S., Riley, M., Schaechter, M., Umbarger, H.E. (Eds.), *Escherichia coli* and *Salmonella*: Cellular and molecular biology. ASM Press, Washington D. C., pp. 721–730.

Raoult, D., Stephane, A., Robert, C., Abergel, C., Renesto, P., Ogata, H., La Scola, B., Suzan, M., Claverie, J.-M., 2004. The 1.2-megabase genome sequence of mimivirus. Science 306, 1344–1350.

Salanoubat, M., Genin, S., Artiguenave, F., Gouzy, J., Mangenot, S., Ariat, M., Billault, A., Brottier, P., Camus, J.C., Cattolico, L., Chandler, M., Choisene, N., Claudel-Renard, S., Cunnac, N., Gaspin, C., Lavie, M., Molsan, A., Robert, C., Saurin, W., Schlex, T., Siguier, P., Thebault, P., Whalen, M., Wincker, P., Levy, M., Weissenbach, J., Boucher, C.A., 2002. Genome sequence of the plant pathogen *Ralstonia solanacearum*. Nature 415, 497–502.

Sambrook, J., Russell, D.W., 2001. Molecular cloning: a laboratory manual, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York N.Y.

Schelle, M.W., Bertozzi, C.R., 2006. Sulfate metabolism in mycobacteria. ChemBioChem 7, 1516–2154.

Serwer, P., Haynes, S.J., Thomas, J.A., Hardies, S.C., 2007. Propagating the missing bacteriophages: a large bacteriophages in new class. Virol. J. 4, 21.

Snider, J., Houry, W.A., 2006. MoxR AAA+ ATPases: a novel family of molecular chaperones? J. Struct. Biol. 156, 200–209.

Stothard, F., Wishart, D.S., 2005. Circular genome visualization and exploration using CGView. Bioinformatics 21, 537–539.

Strott, C.A., 2002. Sulfination and molecular action. Endocr. Rev. 23, 703–732.

Sun, M., Serwer, P., 1997. The conformation of DNA packaged in bacteriophage G. Biophys. J. 72, 958–963.

Suttle, C.A., 2005. Viruses in the sea. Nature 437, 356–361.

Tanaka, H., Negishi, H., Maeda, H., 1990. Control of tobacco bacterial wilt by an avirulent strain of *Pseudomonas solanacearum* M4S and bacteriophage. Ann. Phytopathol. Soc. Jpn. 56, 243–246.

UniProt Consortium, 2007. The Universal Protein Resource (UniProt). Nucleic Acids Res. 35, D193–D197.

Wheeler, D.L., Barrett, T., Benson, D.A., Bryant, S.H., Canese, K., Chetvrnin, V., Church, D.M., DiCuccio, M., Edgar, E., Federhen, S., Geer, L.Y., Kapustin, Y., Khovayko, O., Landsman, D., Lipman, D.J., Madden, T.L., Maglott, D.R., Ostell, J., Miller, V., Pruitt, K.D., Schuler, G.D., Sequeira, E., Sherry, S.T., Sirotkin, K., Souvorov, A., Starchenko, G., Tatusov, R.L., Tatusova, T.A., Wagner, L., Yaschenko, E., 2007. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 35, D5–D12.

Wommack, K.E., Colwell, R.R., 2000. Viroplankton: viruses in aquatic ecosystems. Microbiol. Mo. Biol., Rev. 64, 69–114.

Wozniak, J.A., Zhang, X.-J., Weaver, L.H., Mattheus, B.W., 1994. Structural and genetic analysis of the stability and function of T4 lysozyme. In: Karam, J., Drake, J.W., Kreuzer, K.N., Mosig, G., Hall, D.H., Eiserlig, F.A., Black, L.W., Spicer, E.K., Kutter, E., Carlson, K., Miller, E.S. (Eds.), Molecular biology of bacteriophage T4. ASM press, Washington, D.C., pp. 332–339.

Yabuuchi, E., Kosako, V., Yano, I., Hotta, H., Nishiuchi, Y., 1995. Transfer of two *Burkholderia* and an *Alcaligenes* species to *Ralstonia* gen. nov.: proposal of *Ralstonia picketii* (Ralston, Palleroni and Doudoroff 1973) comb. nov., *Ralstonia solanacearum*

(Smith 1896) comb. nov. and *Ralstonia eutropha* (Davis 1969) comb. nov. Microbiol. Immunol. 39, 897–904.
Yamada, T., Kawasaki, T., Nagata, S., Fujiwara, A., Usami, S., Fujie, M., 2007. New bacteriophages that infect the phytopathogen *Ralstonia solanacearum*. Microbiology 153, 2630–2639.
Yamada, T., Onimatsu, H., Van Etten, J.L., 2006. Chlorella viruses. Adv. Virus Res. 66, 293–336.
Yano, S., Wakayama, M., Tachiki, T., 2006. Cloning and expression of an alpha-1,3-glucanase gene from *Baccilus circulans* KA-304: the enzyme particulates in protoplast formation of *Schizophyllum commune*. Biosci. Biotechnol. Biochem. 70, 1754–1763.